IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	: John C. Hardwick	Art Unit	: 2626
Serial No.	: 10/046,666	Examiner	: Paul V. Harper
Filed	: January 16, 2002	Conf. No.	: 1168
Title	: SPEECH SYNTHESIZER		

MAIL STOP AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REPLY TO ACTION OF FEBRUARY 27, 2007

Applicant requests reconsideration and withdrawal of the current rejections in view of the following remarks.

Claims 1-77 are pending with claims 1 and 38 being independent.

A. Section 101 Rejection

The claims have been rejected under section 101 as being directed to non-statutory subject matter. Applicant requests reconsideration and withdrawal of this rejection because the claims are not directed to a mathematical algorithm in abstract. Rather, the claims are directed to the practical application of the recited signal processing techniques to the processing of digital speech.

The "Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility" ("Interim Guidelines") state, at page 23, that in order to determines that a claimed invention preempts a section 101 judicial exception such as an abstract idea, the Examiner must identify the abstraction and explain why the claim covers every substantial practical application thereof. The Examiner has neither identified an abstraction nor explained why the claim covers every substantial practical application of that abstraction. Moreover, since the claims are limited to the practical application of processing of digital speech, they would not cover applications in other fields such as the processing of digital video or instrumental music. As such, the claims do not preempt a section 101 judicial exception and, therefore, the claims recite patentable subject matter.

In addition to not preempting an abstract idea, the claims recite the useful, tangible and concrete result of producing a set of digital speech samples. In particular, claim 1 recites "combining the first signal samples with the second signal samples to produce a set of digital speech samples corresponding to the selected voicing state" in the context of a method of "synthesizing a set of digital speech samples corresponding to a selected voicing state from speech model parameters." Similarly, claim 38 recites "combining the first signal samples with the second signal samples to produce the digital speech samples for the subframe corresponding to the selected voicing state" in the context of a method of "decoding digital speech samples corresponding to a selected voicing state from a stream of bits."

1. A set of digital speech samples is useful.

Applicant had previously argued that, as evidenced by the industry that has developed around digital speech processing techniques such as are recited in claims 1 and 38, the digital speech samples produced by the methods of claim 1 and 38 are certainly useful. In view of the Examiner's position that applicant has not addressed the issue of tangibility, and the Examiner's not providing any indication that the results are not useful, applicant assumes that the Examiner agrees that the methods of claims 1 and 38 produce useful results.

2. A set of digital speech samples is tangible.

The Interim Guidelines state, at page 21, that the claims must recite a practical application of a technique in order to be tangible. The production of digital speech samples is certainly a practical application of the recited processing techniques. The digital speech samples may be used, for example, by a telephone handset that employs a digital-to-analog converter and a speaker to produce audible speech. However, to require the claims to recite the production of audible speech in order to be directed to patentable subject matter would lead to the absurd result that a handset that performs the recited techniques to produce digital speech samples and then converts the digital speech samples to audible speech would be said to be practicing patentable subject matter while a server that performs the identical techniques but either transmits the digital speech samples to a handset for audible output or stores the digital speech samples for later use would not be said to be practicing patentable subject matter.

3. A set of digital speech samples is concrete.

The Interim Guidelines indicate that a "concrete" result is one that is substantially repeatable. As digital processing techniques are, by their very nature, repeatable, the production of a set of digital speech samples is a concrete result.

Accordingly, for at least these reasons, the claims are directed to statutory subject matter and the rejection under section 101 should be withdrawn.

B. Section 103 Rejection

Claims 1-6, 16, 27, 28, 37-41, 43, 44, 59, 60, 62 and 63 have been rejected as being unpatentable over Griffin (U.S. Patent No. 5,701,390) in view of Barnwell. Claims 7, 42, 45, 46, 49, 61, 64, 65 and 68 have been rejected as being unpatentable over Griffin in view of Barnwell and allegedly well known prior art.

Applicant again requests withdrawal of these rejections for the reasons presented previously. In the interest of completeness, applicant's prior arguments are repeated below with the Examiner's response to those arguments noted in bold, italicized text and addressed in italicized text. As previously argued by applicant:

1. Griffin and Barnwell do not describe or suggest the subject matter of claim 1, which is directed to synthesizing a set of digital speech samples corresponding to a selected voicing state using first and second digital filters computed from first and second frames of speech model parameters.

Claim 1 is directed to a method of synthesizing a set of digital speech samples corresponding to a selected voicing state (e.g., voiced, unvoiced or pulsed) from speech model parameters. The method includes dividing the speech model parameters into frames that include pitch information, voicing information determining the voicing state in one or more frequency regions, and spectral information. First and second digital filters that have frequency responses that correspond to the spectral information in frequency regions where the voicing state equals the selected voicing state are computed using, respectively, first and second frames of speech model parameters. Then, a set of pulse locations are determined and sets of first and second signal samples are produced from the pulse locations and, respectively, the first and second

digital filters. The first signal samples are combined with the second signal samples to produce a set of digital speech samples corresponding to the selected voicing state.

Griffin (U.S. Patent No. 5,701,390), which is commonly assigned with the present application, is directed to a multi-band excitation ("MBE") system that, like claim 1, employs frames of speech model parameters that include pitch information, voicing information, and spectral information. However, Griffin does not describe or suggest the recited computing of first and second digital filters, or the recited use of the digital filters, along with pulse locations, to produce sets of first and second digital samples that are combined to produce a set of digital speech samples.

Applicant recognizes that the rejection notes that "it might be argued that the use of fundamental frequency information determines a set of pulse locations." However, even assuming for sake of argument that this is correct, this in no way changes the fact that Griffin nowhere describes or suggests the use of first and second digital filters, along with pulse locations, to produce sets of first and second digital samples that are combined to produce a set of digital speech samples, as recited in claim 1.

Barnwell, which is a chapter from a textbook on speech coding that describes a pitch-excited linear predictive coder ("LPC"), also fails to describe or suggest the recited computing and use of first and second digital filters.

The rejection indicates that Griffin teaches computing first and second digital filters at Fig. 2 and col. 4, lines 38-65. However, that passage merely mentions that unvoiced frequency band components may be generated from a filter response to a random noise signal, where the filter has a magnitude of approximately the spectral envelope in unvoiced bands and approximately zero in voiced bands. The passage nowhere describes or suggests using the filter in conjunction with pulse locations.

The Examiner responds to this argument by noting that (1) the passage describes the generation of voicing information using regenerated spectral phase information and (2) Barnwell is included to support the use of pulse locations. *As to the Examiner's first point, while applicant agrees that the passage describes the generation of voicing information, such generation of voicing information does not involve computing first and second filters and has nothing to do with the passage's statement that unvoiced frequency band components may be*

generated from a filter response to a random noise signal. As to the Examiner's second point, Barnwell is addressed below.

The final rejection also indicates that Griffin teaches the determining of spectral and voicing information for frequency bands of a frame at the abstract and col. 5, lines 58-62, and that the determining of voicing information necessarily determines pulse excitation locations. This conclusion by the Examiner is not understood. Moreover, even assuming for sake of argument that it is correct, it would not lead to the recited use of digital filters in conjunction with the pulse locations since, as noted above, Griffin states that the filter response is to a random noise signal.

The Examiner responds to this by arguing that (1) Barnwell describes the relationship between fundamental frequency and pitch, (2) Barnwell describes how a train of pitch pulses can be used to excite a digital filter to produce a voiced signal, (3) Griffin teaches that fundamental frequency information is used (not just random noise), and (4) Barnwell describes a pulse generator that generates pulses corresponding to voiced speech and a noise generator that generates a random noise signal corresponding to unvoiced speech. *As to the Examiner's third point, as noted above, while Griffin describes the use of fundamental frequency information, Griffin does not describe the use of this information in conjunction with Griffin's use of a filter response to a random noise signal to generate unvoiced frequency components.*

As to the Examiner's first, second and fourth points, even assuming for sake of argument that the Examiner's characterization of Barnwell is correct, this in no way remedies the failure of Griffith, Barnwell and their combination to describe or suggest the use of first and second digital filters, along with pulse locations, to produce sets of first and second digital samples that are combined to produce a set of digital speech samples, as recited in claim 1.

Recognizing that Griffin does not describe or suggest determining a set of pulse locations, producing sets of first and second signal samples using the digital filters and the pulse locations, and combining the first and second signal samples to produce digital speech samples, the rejection asserts that doing so was well known, as evidenced by Barnwell. Applicant notes that the Examiner states:

> Barnwell illustrates (clarifies) the connection between the fundamental frequency (as taught by Griffin) and pulse locations as claimed when used to excite a filter

> (programmed with spectral information) during a voiced state. Barnwell also illustrates the sequential nature of the process: a first set of spectral coefficients program the first digital filter and when excited produce the first set of digital samples; the second set of spectral coefficients program the second filter and when excited produce the second set of digital samples, etc. These outputs are combined to produce the reconstituted digital signal.

Applicant has reviewed Barnwell and does not see where Barnwell sets forth the noted illustration and, to the extent that the Examiner continues to maintain that such illustration may be found in Barnwell, applicant requests that the Examiner provide an explanation of where it can be found.

The Examiner notes that Barnwell, at Fig. 5.2, page 88, describes the input of pitch information to a pulse generator which for voice signals excites a filter (linear predictor) which is configured with spectral information (LPC Coefficients). *Even assuming for sake of argument that the Examiner's characterization of Barnwell is correct, this in no way describes or suggests the use of first and second digital filters, along with pulse locations, to produce sets of first and second digital samples that are combined to produce a set of digital speech samples, as recited in claim 1, and would in no way have led one of ordinary skill in the art to modify Griffin to do so.*

Moreover, even assuming for sake of argument that Barnwell somehow illustrates the points noted by the Examiner, this seems to simply be a repeat of the Examiner's argument in the previous rejection, where the Examiner stated:

> Barnwell teaches the more specific operations of using voicing information along with spectral information (or filter coefficients) to produce the synthesized output (i.e., pulse generator with pitch locations exciting the filter). When Barnwell's teaching are combined with those of Griffin you get "producing of sets of first and second signal samples using the digital filters and pulse locations", and "the recited combining of the first and second signal samples to produce digital speech samples."

As previously noted, applicant strongly disagrees. First, the passage of Barnwell identified in the rejection (pages 85-89) merely describes well known LPC techniques and in no way describes or suggests the recited producing of sets of first and second signal samples using the digital filters and the pulse locations, or the recited combining of the first and second signal samples to

produce digital speech samples. Accordingly, for at least these reasons, the rejection of claim 1 and its dependent claims should be withdrawn.

The Examiner responds to this argument by stating that (1) Griffin teaches the generation of synthetic speech with the input of fundamental frequency and spectral (coefficient) information where a filter is defined by the coefficients used to program it (Fig. 2), (2) that, since each frame corresponds to spectral information, sequential frames will define sequential filters (hence a first and second filter), and (3) that Barnwell further clarifies the connection between pulse locations (and fundamental frequency) and the excitation of a digital filter. *As to the Examiner's first point, and as discussed above, Griffin does not describe the use of a filter in the manner argued by the Examiner. As to the Examiner's second and third points, under the Examiner's own logic, if sequential frames could be said to have different filters as a result of their having different spectral information, they would also have different pulse locations as a result of having different fundamental frequencies, such that the different filters would not be used in conjunction with the same pulse locations to produces sets of first and second digital samples.*

<u>2. There would have been no motivation to combine Griffin and Barnwell in the manner set forth in the rejection, since Griffin is directed to MBE coder, and Barnwell is directed to a LPC coder, which is a substantially different class of coder.</u>

Griffin and Barnwell are directed to different classes of coders. As such, nothing in Barnwell's description of a LPC coder would have led one of ordinary skill in the art to modify Griffin's MBE coder to produce a coder such as is recited in the claims. Moreover, the rejection does not identify any such motivation. Rather, the rejection merely asserts that it would have been obvious to do so because Barnwell allegedly describes the features missing from Griffin.

The Examiner responds to this argument by stating that Barnwell was included because it teaches well known techniques that can be used in data compression and it clarifies the connection between the fundamental frequency and pulse locations and the programming of a filter with spectral information. *Even assuming for sake of argument that the Examiner's characterization of Barnwell is correct, Barnwell's teaching of known techniques and any other*

clarification offered by Barnwell would not have provided any motivation for one of ordinary skill in the art to modify Griffin.

While the argument by the Examiner might be said to assert that the motivation to combine the references would come from a desire to reduce the bandwidth required by Griffin's system, there is no indication that such a reduction would result. Indeed, as Griffin's system is already directed to using a low bandwidth (3.6 kbps) system (see col. 5, lines 60-63), it seems likely that attempting to incorporate Barnwell's substantially different approach would result in an increase in the bandwidth requirement.

3. Griffin and Barnwell do not describe or suggest the subject matter of claim 38, which is directed to decoding a stream of bits to produce speech samples corresponding to a subframe by computing impulse responses for the subframe and a previous subframe, and applying pulse locations for the subframe to produce sets of first and second signal samples that are combined to produce the speech samples.

Claim 38 is directed to decoding digital speech samples corresponding to a selected voicing state from a stream of bits. The stream of bits is divided into a sequence of frames that each contain one or more subframes. Speech model parameters are decoded from the stream of bits for each subframe in a frame, with the decoded speech model parameters including at least pitch information, voicing state information and spectral information. A first impulse response is computed from the decoded speech model parameters for a subframe, and a second impulse response is computed from the decoded speech model parameters for a previous subframe. Thereafter, a set of pulse locations is computed for the subframe, and sets of first and second signal samples are produced from the pulse locations and, respectively, the first and second impulse responses.

Griffin and Barnwell fail to describe or suggest the subject matter of claim 38 for the reasons discussed above with respect to claim 1. In addition, neither Griffin nor Barnwell anywhere describes or suggests applying pulse locations for a subframe to an impulse response computed using decoded speech model parameters for the subframe and decoded speech model

parameters for a previous subframe. Nor does the rejection provide any indication of where such application may be found in Griffin or Barnwell.

Accordingly, applicant submits that all claims are in condition for allowance.

No fee is believed to be due. Please apply any charges or credits to Deposit Account No. 06-1050.

Respectfully submitted,

Date: May 9, 2007

John F. Hayden
Reg. No. 37,640

Fish & Richardson P.C.
1425 K Street, N.W.
11th Floor
Washington, DC 20005-3500
Telephone: (202) 783-5070
Facsimile: (202) 783-2331

40414421.doc